Virtual Center for Network and Security Data

University of Michigan
Merit Network
University of Wisconsin
University of Washington
Internet2

merit
NETWORK INC

PREDICT Workshop
September 27, 2005
Newport Beach, CA

Participants

- Phase I
  - Farnam Jahanian, University of Michigan
  - Morley Mao, BEACON
  - Manish Karir, Merit Network
- Phase II
  - Paul Barford, University of Wisconsin
  - Dave Dittrich, University of Washington
  - Matthew Zekauskas and Rick Summerhill, Internet2

Blackhole Datasets

- A Blackhole/Dark IP/Telescope sensor monitors an unused globally advertised address block that contains ***no active hosts.*** Traffic is the result of ***DDoS backscatter***, ***worm propagation***, ***misconfiguration***, or other ***scanning***
- Long-term Trend Datasets
  - One year dataset
  - Three /24 sensors
  - Counts of
    - the number of packets,
    - unique source Ips
    - unique MD5 payload checksums for each port
- Raw Sample Datasets
  - Include all the raw packets captured over a short period
    - All of April 2004.
  - Quite large
  - These datasets are also in anonymized pcap format.

Blackhole Event Specific Datasets

- The format of the data is pcap, one file per day.
- Filtered to only include specific ports.
- The source and destination addresses are anonymized
  - prefix preserving on all 32-bits with payloads md5'd
  - last 8 bit constant substitution with full payload also available

- Witty:
  - UDP src port 4000
  - March 18, 2004 to March 25, 2004
- MySQLbot worm:
  - TCP dst port 3306
  - Jan 25, 2005 to Feb 1, 2005
- Sasser
  - TCP dst port 445
  - April 29, 2004 to May 6, 2004
- Dabber
  - TCP dst port 5554
  - May 9, 2004 to May 29, 2004
- TCP/42 WINS
  - TCP dst port 42
  - Dec 8, 2004 to Jan 10, 2005
- TCP/6101 Veritas
  - TCP dst port 6101
  - Jan 10 2005 to Jan 17, 2005
- Bagle/MyDoom backdoor
  - TCP dst port 2745 and 3127,
  - March 20, 2004 to March 27, 2004

BEACON Datasets

- Beacons routing infrastructure. This infrastructure is where the routing prefixes changes are injected into the Internet. Currently, there are four sites that host the Beacons
- Hourly tar.gz files with
  - BGP routing data associated with the changes in BEACON prefixes (MRT format)
  - UDP packet traces including loss, delay, reordering, jitter, and throughput. (Plaintext)

Merit Datasets

- MichNet is Merit's regional research and education network, is the largest IP network in Michigan
- Collection
  - Collection at MichNet's 4 peering routers
  - Files tar.gz'd in various frequencies
  - May 1st-present
- Netflow
  - Sampling varies from 1:1 - 1:1000
  - Hourly files
  - Netflow version 5
  - Stored in flow-tools format
  - SRC and DST have last 8 bits anonymized (constant substitution)
- BGP Routing data
  - Table dumps every four hours
  - Updates in 15 minute files
  - MRT format

Upcoming datasets

- Broader coverage
  - Netflow
    - University of Wisconsin border router
    - University of Washington border router
    - Internet2 backbone
  - BGP
    - University of Wisconsin border router
    - University of Washington border router
  - Blackhole
    - iSink (University of Wisconsin)
- New datasets
  - Honeypot data (University of Washington)
  - DSHIELD Intrusion Detection data (University of Wisconsin)
  - Internal routing data (Internet2)
- Security event correlation datasets
  - Similar to blackhole data; event specific and “normal” traces
  - Blackhole data, routing data, netflow data, honeypot data together